==========

1 [main text]

One can get a physical sense of the importance of the extended gene environment by looking at a situation where a genomic variant rearranges the extended gene structure without affecting coding regions. We found such an example in the breast cancer cell line T47D, where a 130Kb heterozygous deletion changes the chromosome to link a distal enhancer to the promoter and results in the activation of the well-known oncogene ERBB4 (Fig. 2F). (The enhancer not being connected ERBB4 in normal breast tissue.) This heterozygous deletion is located around 45Kb downstream from the ERBB4 promoter region and potentially merges two Hi-C TADs in an allele-specific way. We tested this hypothesis through CRISPR editing, by excising an 86bp sequence that contains the CTCF binding sites at the boundary of the two Hi-C TADs from the wild-type allele in T47D cells. This excision confirmed the elevated ERBB4 expression upon CRISPR deletion (as measured by PCR).

- expt type: CRISPR editing and PCR-based expression quantification; CRISPR-engineered deletion to evaluate the effect of an SV on enhancer linkages
- what it demonstrates in paper: modifying an extended gene neighborhood (without directly changing the coding region) can result in changes in gene expression through the introductin of enhancer-promoter interactions (in this case, elevating the expressio of ERBB4 in T47D cells)
- associated figure(s): Fig. 2F,E
- sections (in supp): (not in original supp -- included in new version?)
- primary lit (if any): --
- data files (and their IDs): --

==========

2 [main text]

As expected, we found that the target genes of MYC are significantly up-regulated in numerous cancer types -- in fact, the most up-regulated of any TF -- consistent with its well-known role as an oncogenic TF. We further validated MYC's regulatory effects using knockdown experiments in breast cancer (Fig. 3). Consistent with our predictions, the expression of MYC targets is significantly reduced after MYC knockdown in MCF-7 (Fig. 3B).

- expt type: gene knockdowns followed expression quantifcation
- what it demonstrates in paper: The regulatory network that we construct recapitulates intuitive biolgy: Genes identified as targets of MYC exibit reduced expression upon MYC knockdown. (Our models posits that MYC binds to the promoter of target genes to upregulate expression)
- associated figure(s): Fig. 3b
- sections (in supp): --
- primary lit (if any): References 4 & 5
- data files (and their IDs): --

==========

3 [main text]

As an RBP, SUB1 has not previously been associated with cancer, so we sought to investigate its role. Knocking down SUB1 in HepG2 cells significantly down-regulated its targets (Fig. 3D), and the decay rate of SUB1 targets is lower than those of non-targets (see suppl.). Moreover, we found that up-regulation of SUB1 targets may lead to decreased patient survival in some cancer types (Fig. 3D).

- expt type: gene knockdowns followed expression quantifcation
- what it demonstrates in paper: SUB1 knockdown in HepG2 results in diminished expression of its target genes (oncogenes). (Our models posits that SUB1 stabilizes RNA tranascripts of 3UTRs of targets, thereby increasing transcript lifetime in the cell).
- associated figure(s): Fig. 3D
- sections (in supp): --
- primary lit (if any): --
- data files (and their IDs): --

{{linked to 2 and 3}} [main text]

Similarly, with respect to RBPs, we found that the top co-regulatory partner of SUB1 is MYC (SUB1 is a direct target of MYC in many cell types, see suppl. sect.). SUB1 and MYC together form many FFLs in the regulatory network. We hypothesized that MYC can bind to the promoter regions of key oncogenes to initiate their transcription, whereas SUB1 binds to 3UTRs to stabilize oncogenes at the level of RNA transcripts. Such synergistic collaboration between MYC and SUB1 results in overexpression of several key oncogenes and leads to proliferation of cancer cells (see suppl. sect. xxx). To validate this hypothesis, we knocked down MYC and SUB1 separately in HepG2 and used qPCR to quantify changes in gene expression. As expected, the expression of oncogenes (such as MCM7, BIRC5, and ATAD3A) is significantly reduced (Fig. 3E).

To demonstrate the utility of our ENCODE resource, we instantiated our prioritization workflow in a few select cancers and experimentally validated the results. In particular, as described above, we subjected some key regulators, such as MYC and SUB1, to knockdown experiments (Fig. 3B and 3D) and we measured the effect of SVs on enhancer linkages via CRISPR engineered deletions (Fig. 2E).

- expt type:
- what it demonstrates in paper:
- associated figure(s):
- sections (in supp):
- primary lit (if any):
- data files (and their IDs):

==========

4 [main text]

Finally, we selected key SNVs based on their disruption of enhancers with strong influence on gene expression. These SNVs were prioritized based on mutation recurrence in breast-cancer cohorts, as well as enhancer motif disruption scores. Of the eight motif-disrupting SNVs that we tested, six exhibited consistent up- or down-regulation relative to the wild-type in multiple biological replicates.

One particularly interesting example is in an intronic region of CDH26 in chromosome 20 (Fig. 6C). The signal shapes for both histone modification and chromatin accessibility (DNase-seq) data indicate its active regulatory role as an enhancer in MCF-7. This was further confirmed by STARR-seq (Fig. 6C). Hi-C and ChIA-PET linkages indicated that the region is within a topologically associated domain (i.e., a "TAD") and validated a regulatory connection to the breast-cancer-associated gene SYCP2. We further observed strong binding of many TFs in this region in MCF-7. Motif analysis predicts that a common mutation in breast cancer affects this region, and significantly disrupts the local binding affinity of several TFs, such as FOSL2 (Fig. 6C).

- expt type: introducing SNVs followed by luciferase assays to measure gene expression
- what it demonstrates in paper: a) We correctly prioritize SNVs (using recurrence analysis and scoring motif disruptions) that may disrupt gene expression through a mechanism by which enhancers are impaired; b) our predicted enhancers are consistent with these expression changes
- associated figure(s): Fig. 6c
- sections (in supp): --
- primary lit (if any): --
- data files (and their IDs): --